

System For Reducing Server Loading During Content Delivery

Kaushik Barde
Richard Hartman
Gopinath Rebala

System For Reducing Server Loading During Content Delivery

Technical Field

[0001] The present invention relates generally to electronic multimedia exchange, and, more particularly, to reducing server load during content delivery.

Background

[0002] The World Wide Web (WWW) is a system of Internet servers that support specially formatted documents. The documents are formatted in a script called HTML (HyperText Markup Language) that supports links to anything that is capable of being transmitted digitally over the Internet, such as documents, graphics, audio files, video files and the like. There are several applications called "web browsers" that simplify access to the WWW. Two notable examples are Netscape Navigator™ and Microsoft's Internet Explorer™.

[0003] For a transaction on the WWW between a client terminal and a host server, the client terminal retrieves a web object from the host server. The host server is identified through a Uniform Resource Locator (URL), which is the global address of documents and other resources on the WWW. The client terminal normally accesses the Internet through an Internet Access Service Provider (IASP). A cache, located within the IASP network, functions as an intermediary in transactions involving retrieval of objects from servers by a client terminal. The cache within the IASP saves a copy of a retrieved object for itself when the object is moved from the server to the requesting client terminal.. This caching operation is transparent to the user and, under normal circumstances, does not result in substantial delay from the copying operation. The copying operation is performed simultaneously with the retrieval from the server and the delivery to the client terminal.

[0004] The cache within the IASP network satisfies subsequent requests for objects that are stored therein, thereby obviating the necessity of retrieving the object from the original server. This reduces the delay as perceived by the user to access the object and saves bandwidth on links that connect the IASP network to the Internet. HTTP is the underlying protocol used by the World Wide Web. HTTP defines how messages are formatted and transmitted and what actions web servers and browsers should take in response to various commands. Direct browsing by multiple clients to a particular site using HTTP tends to result in overloading of the host.

[0005] A further problem currently exists with satisfying subsequent requests for an object from the cache. The copy of the object stored in the cache may differ from the object in the server if the latter is modified after the initial request for the object. When this occurs, the copy of the object provided to the requesting client from the cache may not be current.

[0006] The disadvantages associated with current, server download techniques have made it apparent that a new technique for downloading objects is needed. Given a host server web page, the new technique should allow multiple clients to rapidly and efficiently download the entire page. The new technique should also update cached information to minimize outdated information transfers. The present invention is directed to these ends.

Summary of the Invention

[0007] It is an object of the present invention to provide a system for allowing multiple clients to rapidly and efficiently download entire web pages. It is another object of the present invention to update cached information to minimize outdated information transfers.

[0008] In accordance with the present invention, a server load reduction system is disclosed. The server load reduction system includes a master URL

containing data. The system further includes a proxy browser, which conducts a browse operation to request the data contained in the master URL. This browse operation is conducted through a proxy server. The proxy server is capable of receiving the data from the master URL. The proxy server includes logic operative to record and distribute the data to a client server. Logic contained in the proxy browser is operative to notify a client server to load the data when the proxy server contains all of the data.

[0009] Additional advantages and features of the present invention will become apparent from the description that follows and may be realized by the instrumentalities and combinations particularly pointed out in the appended claims, taken in conjunction with the accompanying drawings.

Brief Description Of The Drawings

[0010] For a more complete understanding of the invention, there will now be described some embodiments thereof, given by way of example, reference being made to the accompanying drawings, in which:

[0011] FIGURE 1 is a perspective view of a server load reduction system in accordance with an embodiment of the present invention;

[0012] FIGURE 2 is a schematic diagram of a server load reduction system in accordance with an embodiment of the present invention; and

[0013] FIGURE 3 is a block diagram of a server load reduction system in accordance with an embodiment of the present invention.

Detailed Description Of The Preferred Embodiment

[0014] The present invention is illustrated with respect to a server load reduction system, particularly suited to electronic multimedia transfer and storage. However, the present invention is applicable to various other uses that

may require reduction of server loads, as will be understood by persons skilled in the art.

[0015] Referring to FIGURE 1, a perspective view of a server load reduction system 10, in accordance with one embodiment of the present invention, is illustrated. The system 10 includes a group of users 12 from the same client operating a group of user terminals 13 and downloading the same master URL 14 from the WWW 16, as will be understood by one skilled in the art. The current invention is browser type independent. In other words, one of the users may run Netscape Navigator™ while another runs Internet Explorer™. After the client group of users 12 requests the master URL 14 through the group of user terminals 13, the client system 18 activates and downloads the data from the master URL 14 for storage and use in the group of user terminals 13, as will be discussed later.

[0016] Referring to FIGURE 2, a schematic diagram of a server load reduction system 10, in accordance with one embodiment of the present invention, is illustrated. The load reduction system 10 includes a master Uniform Resource Locator (URL) 14. The URL is the global address of a set of documents and other resources on the WWW 16. When a group of computer users 12 for a specific client browse to a particular HTTP site, that site is subjected to large amounts of loading proportional to the number of clients attempting access. For the current embodiment, one client downloads the entire master URL 14 and then accesses the client proxy server 28 rather than the master URL 14, thus freeing the master URL 14 for other clients.

[0017] This system 10 operates ideally with both unicast and multicast sections. The unicast section includes a proxy browser 30, a proxy server 28, a client server 32 (first client server), a keypad manager 34 and a control message 36 for the client group of user terminals 13. The unicast section transmits a

portion of the data in the master URL 14 to the group of user terminals 13, as will be discussed later.

[0018] The multicast section includes the proxy browser 30, the proxy server 28, the client server 32, a multicast server 38, a multicast client server 40 (second client server), a multicast client storage location 41 with a browser cache 42, and a connection 43 to the group of user terminals 13. The multicast section transmits the entire contents of the master URL 14 to the group of user terminals 13, as will be discussed later.

[0019] As previously mentioned, the system 10 includes a proxy browser 30, which facilitates WWW searching. The proxy browser 30 functions as follows: the client initiates a browse command on the client server 32, which sends a command to the proxy browser 30 to load a particular URL (e.g. www.xyz.com), which subsequently sends a URL request (browse operation) 44 through the proxy server 28.

[0020] The proxy browser 30 conducts a first browse operation to request a portion of the data contained in the master URL 14. The proxy browser 30 then sends a unicast signal containing the portion of the master URL 14 through the client server 32 and to the keypad manager 34, which funnels control messages to the client terminal 13, as will be understood by one skilled in the art. The client terminal 13 then views the portion of the master URL 14 and decides whether to download the entire master URL 14. If the client affirmatively signals the proxy browser 30 to download the entire master URL 14, the multicast section activates and the proxy browser 30 downloads the data contained in the master URL to 14 the proxy server 28.

[0021] The proxy browser 30 contains logic operative to notify a multicast server 38 to load the data when the proxy server 28 contains all of the master URL data. In other words, the multicast server 38 loads the specified URL when

the proxy server 28 has received and stored all of the URLs and data associated with the master URL 14 and when the client has received a command from the proxy browser 30 to load the master URL. The data transfer is ideally conducted through SERGE, which is a reliable information transfer system for multicast systems, as will be understood by one skilled in the art. The multicast server then sends the data to the multicast client storage location 41 through multicast channels 46, as will be understood by one skilled in the art. The browser cache 42, attached to the multicast client storage location 41, receives and stores the master URL data for access and use by the group of computer users 12.

[0022] As previously mentioned, the system 10 also includes a proxy server 28. The embodied proxy server 28 has an internal proxy server cache 29 in which it stores files for future requests from the proxy browser 30. The proxy server 28 contains logic operative to record the master URL data. The proxy server 28 also contains logic operative to signal the proxy server cache 29 when a portion of the master URL 14 is modified. When signaled, the proxy server 28 reloads the data contained in the master URL 14. The proxy server 28 also includes a distribution mechanism, which distributes the data to the client server 32.

[0023] Referring to FIGURE 3 in view of FIGURES 1 and 2, a block diagram of a server load reduction system 10, in accordance with one embodiment of the present invention, is illustrated. The logic starts in operation block 52 when the client conducts a browse operation (through the proxy browser 30) to find a master URL 14.

[0024] A determination is then made in inquiry block 54 whether the URL is the desired master URL. For a negative response, the logic flows again through operation block 52 until a positive response is obtained.

[0025] For a positive response, operation block 56 activates; and the proxy browser 30 requests the data stored in the master URL 14. Subsequently, operation block 58 activates, and the proxy server 28 receives the master URL data.

[0026] A determination is then made in inquiry block 60 whether the proxy server 28 contains all the master URL data. For a negative response, operation block 58 reactivates; and the proxy server 28 continues to receive master URL data.

[0027] Otherwise, operation block 62 activates, and the proxy browser 30 notifies the client server 32 to download the data from the proxy server.

[0028] In operation, in view of FIGURES 1,2 and 3, the client conducts a browse operation with the proxy browser 30 to load a desired master URL 14. When a potential master URL is found, the unicast portion of the system activates and a portion of the master URL 14 is stored in the proxy server 28. When the portion of the master URL 14 is completely stored in the proxy server 28, the proxy server 28 notifies the client server 32. The client is then free to view the portion of the master URL without requiring multicast operations.

[0029] When the client decides that the potential master URL is the desired master URL 14, the multicast portion of the system activates. Subsequently, the proxy browser 30 requests all the data contained in the master URL 14, and that data is received in the proxy server 28. The proxy server 28 then sends the data via multicast lines 48 to the multicast server 38 which stores the data and notifies the multicast client server 40 when the multicast server 38 contains all of the data in the master URL 14. The multicast client server 40 then receives the data through multicast signal lines 46 and stores the master URL data in a browser cache 42. The group of user terminals 13, associated with the

client, then load portions of the master URL 14 as desired by the group of computer users 12.

[0030] It is to be understood that the multicast and unicast portions of the invention may function independently or jointly, as embodied, depending on the amount of information in the master URL 14. For small quantities of data, the unicast section sufficiently transfers the data to the group of user terminals 13. For larger quantities of data, the multicast section provides a greater number of open data transfer channels and therefore substantially increases data flow.

[0031] From the foregoing, it can be seen that there has been brought to the art a new and improved server load reduction system 10. It is to be understood that the preceding description of the preferred embodiment is merely illustrative of some of the many specific embodiments that represent applications of the principles of the present invention. Numerous and other arrangements would be evident to those skilled in the art without departing from the scope of the invention as defined by the following claims: